STUDIA INFORMATICA 2002
Volume 23 Number 2B (49)

Zbigniew OMIOTEK
Wyższa Szkoła Zarządzania i Administracji w Zamościu

Franciszek GRABOWSKI
Politechnika Rzeszowska

POMIARY WYBRANYCH PARAMETRÓW RUCHU W SIECIACH ROZLEGŁYCH

Streszczenie. W artykule zostało przybliżone pojęcie samopodobieństwa w odniesieniu do ruchu obserwowanego w sieciach komputerowych. Zaprezentowano wybrane metody testowania samopodobnego charakteru procesów. Przedstawiono również wyniki pomiarów szybkości wybranych tras ruchu w sieciach WAN. Następnie, wykorzystując metody graficzne, dokonano analizy tych wyników pod kątem występowania samopodobieństwa.

MEASUREMENTS OF SELECTED TRAFFIC PARAMETERS OF WAN'S

Summary. In this paper the idea of self-similarity in relation to the network traffic has been explained. Selected methods of testing self-similarity processes have been presented. The results of the speed selected routes in WAN have been showed as well. Next, taking advantage of graphical methods, these results have been analysed with regard to the idea of self-similarity.

1. Wstęp

W ostatnich latach przeprowadzono szereg badań dotyczących ruchu w sieciach LAN i WAN [1, 2, 3, 4]. Wykazały one, iż właściwości ruchu rzeczywistego znacznie różnią się od tych które wynikają ze stosowanych dotychczas klasycznych modeli ruchu, uwzględniających np. rozkład Poissona. Dotychczas ruch był charakteryzowany zależnościami krótkoterminowymi, dla których funkcja autokorelacji jest postaci wykładniczej. Jednak w rzeczywistości w sieci występują procesy z zależnościami

długoterminowymi z hiperboliczną postacią funkcji autokorelacji. W rzeczywistym ruchu występuje zjawisko zwane wybuchowością, przy czym jest ono obserwowane również dla ruchu zagregowanego w większej skali czasu. Proces wykazujący taką właściwość określa się jako samopodobny. Ogólnie, pojęcie samopodobieństwa oznacza niezmienność właściwości statystycznych procesu przy zmianie skali czasu obserwacji. W przypadku ruchu sztucznego, generowanego przez modele tradycyjne, agregacja powoduje zanik zjawiska wybuchowości.

Wiele badań wykazało znaczący wpływ zależności długoterminowych na takie parametry sieci, jak przepustowość, współczynnik straty pakietów, czas odpowiedzi, zajętość buforów [5]. Wynika to z poważnego obniżenia wydajności kolejkowania w sieciach z zależnościami długoterminowymi. Okazuje się, że rozkład długości kolejek w takich sieciach opada wolniej niż eksponencjalnie [6, 7]. Oznacza to, że pakiet zanim dotrze do miejsca przeznaczenia doznaje w rzeczywistości większego średniego opóźnienia, niż to przewidziane przez modele tradycyjne, takie jak model Markowa. Wnioski wyciągnięte z tych badań mają doniosłe znaczenie dla projektowania protokołów i algorytmów kontroli zatłoczeń [7]. Proponowane są również nowe modele źródeł ruchu, które w większym stopniu uwzględniają rzeczywiste właściwości ruchu sieciowego [8, 9, 10, 11].

Głównym celem tej pracy jest eksperymentalne potwierdzenie tezy, że czas transmisji pakietów w obie strony (ang. *Round-Trip Time*) jest procesem z zależnością długoterminową [12, 13]. Wiele aplikacji wykorzystujących Internet, np. do przesyłania danych audio lub video, jest czułych na wydajność całego Internetu, a parametr *RTT* jest ważną miarą tej wydajności. Badanie rozkładu *RTT* może dać interesujące spostrzeżenia na temat wydajności kolejkowania pakietów wzdłuż ich trasy.

W dalszej części pracy, w rozdziale 2 scharakteryzowano sposób pomiaru opóźnienia pakietów. Następnie w rozdziale 3 zaprezentowano metody badania samopodobieństwa wykorzystane w analizie statystycznej oraz przedstawiono wyniki analizy danych pomiarowych. W rozdziale 4 przedstawiono wnioski wyciągnięte na podstawie tej pracy i krótko przedstawiono plany dalszych badań.

2. Pomiar opóźnienia pakietów

Wartość parametru *RTT* jest określona przez sumę opóźnień, jakich doznaje pakiet w każdej podsieci, przez którą jest przesyłany. Opóźnienie wnoszone przez każde połączenie jest sumą opóźnień przetwarzania, kolejkowania, transmisji i propagacji. Przy stałej długości pakietu i jego trasy na zmianę opóźnienia wpływają jedynie kolejki, których długość zmienia się zależnie od fluktuacji natężenia ruchu. Przy założeniu stałej trasy, opóźnienie pakietu $T(t)$ jest zmienną losową w czasie t. Zmienna $T(t)$ opisuje proces opóźnienia pakietu.

Przedmiotem badań był szereg czasowy T_i otrzymany przez próbkowanie zmiennej $T(t)$ w czasie t. Szereg T_i składał się z próbek procesu $T(t)$ dla $t = t_i$, gdzie $i = 1,2,3,...$. Wartość t_i określała czas wysyłania pakietów kontrolnych. Po odebraniu pakietów odpowiedzi efektem była próbka zmiennej losowej $T(t)$ dla danej trasy.

Podczas badań eksperymentalnych zostały wykonane pomiary parametru *RTT* pakietów przesyłanych pomiędzy dwoma punktami w sieci WAN. Stacja robocza pracująca w sieci komputerowej Wyższej Szkoły Zarządzania i Administracji w Zamościu (*rh.wszia.edu.pl*) wysyłała w odstępach czasu co 5 sekund serie trzech pakietów kontrolnych do wybranych serwerów. Na podstawie trzech zmierzonych czasów odpowiedzi została obliczona i zarejestrowana średnia wartość parametru *RTT* dla danego połączenia. Powyższa procedura była powtarzana co 5 sekund przez 8 dni w terminie od 09.01.2002 r. do 17.01.2002 r. Do celów pomiarowych zostały wykorzystane trzy serwery WWW posiadające różną lokalizację geograficzną. Pierwszy z nich należał do koncernu chemicznego BACHEM w Szwajcarii (*www.bachem.com*), drugi do Uniwersytetu Queen's w Kanadzie (*knot.queensu.ca*), a trzeci do Narodowej Organizacji Turystyki w Japonii (*www.jnto.go.jp*). W wyniku pomiarów zostały otrzymane szeregi czasowe o długości odpowiednio 90850, 40789 i 59663 próbek. Ze względu na ograniczenia oprogramowania zastosowanego w analizie statystycznej z pierwszego i trzeciego szeregu czasowego wykorzystano 50 000 próbek.

3. Analiza statystyczna wyniku pomiarów

Najpopularniejsze metody stosowane do badania samopodobieństwa w szeregach czasowych to: analiza przeskalowanego zakresu, analiza wariancji, analiza oparta na periodogramie, estymator Whittle'a oraz analiza Wavelet [8].

Poziom samopodobieństwa jest określony przez tzw. parametr Hursta[1] (H). Istnieją trzy klasy wartości tego parametru:

$H = 0.5$ – szereg odzwierciedla proces stochastyczny (ruchy Browna);

$0 \le H < 0.5$ – szereg nazywa się „powracający do średniej", jeżeli w danym okresie występuje tendencja wzrostowa, to jest prawdopodobne, że w następnym okresie wystąpi tendencja zniżkowa i na odwrót;

$0.5 < H \le 1$ – szereg posiada własności wzmacniające trend. W tym obszarze obserwuje się fraktalne własności szeregu. Im bardziej wartość H jest zbliżona do 1, tym silniejsze samopodobieństwo charakteryzuje badany proces.

[1] H. E. Hurst był hydrologiem pracującym przez wiele lat przy projekcie zapory na Nilu. Wspólnie z Mandelbrotem badał wahania poziomu rzek. Jest autorem analizy R/S.

W dalszej części tego rozdziału zostanie bliżej scharakteryzowana analiza przeskalowanego zakresu, analiza wariancji i analiza oparta na periodogramie, ponieważ metody te zostały wykorzystane do analizy danych eksperymentalnych w niniejszej pracy.

3.1. Metody badania samopodobieństwa

3.1.1. Analiza przeskalowanego zakresu (analiza R/S)

Dla danego zbioru obserwacji $X = \{X_m, m \in Z^+\}$ ze średnią $\overline{X}(m)$, wariancją $S^2(m)$ i zakresem $R(m)$, statystyka *R/S* jest następująca

$$\frac{R(m)}{S(m)} = \frac{\max(0, \Delta_1, \Delta_2, \ldots, \Delta_m) - \min(0, \Delta_1, \Delta_2, \ldots, \Delta_m)}{S(m)}, \tag{1}$$

gdzie

$$\Delta_k = \sum_{i=1}^{k} X_i - k\overline{X} \text{ dla } k = 1, 2, \ldots, m. \tag{2}$$

Zaobserwowano, iż dla wielu zjawisk naturalnych

$$E\left[\frac{R(m)}{S(m)}\right] \approx cm^H \tag{3}$$

w przypadku gdy $m \to \infty$ oraz c jest stałą dodatnią niezależną od m. Logarytmując dwustronnie wyrażenie (3) otrzymujemy

$$\log\left\{E\left[\frac{R(m)}{S(m)}\right]\right\} \approx H\log(m) + \log(c) \tag{4}$$

dla $m \to \infty$.

Rysując zależność $\log\left\{E\left[\frac{R(m)}{S(m)}\right]\right\}$ w funkcji $\log(m)$ i dopasowując do niej linię prostą można oszacować wartość parametru Hursta. Powyższa wartość będzie estymowana przez nachylenie otrzymanej prostej.

3.1.2. Analiza wariancji

Analiza oparta na wykresie wariancji (ang. *Variance-time plot*) wykorzystuje właściwość procesu samopodobnego polegającą na powolnym opadaniu wariancji w funkcji czasu. Wariancja takiego procesu dla $0 < \beta < 1$ jest wyrażona następująco

$$Var\left[X^{(m)}\right] = Var[X] \cdot m^{-\beta}. \tag{5}$$

Równanie (5) może być zapisane jako

$$\log\left\{Var\left[X^{(m)}\right]\right\} \approx \log[Var(X)] - \beta\log(m). \tag{6}$$

Oszacowanie wartości parametru Hursta polega na wykreśleniu w skali *log-log* zależności $Var(X)$ w funkcji m i dopasowaniu do niej linii prostej o nachyleniu $-\beta$. W procesie dopasowania pomijane są początkowe wartości m. W przypadku zależności krótkoterminowej lub danych niezależnych prostą graniczną będzie prosta o nachyleniu równym -1. Natomiast w przypadku zależności długoterminowej dane ułożą się w prostą o mniejszym (co do wartości bezwzględnej) współczynniku nachylenia. Oszacowanie wartości β pozwala obliczyć parametr Hursta z następującego wzoru

$$H = 1 - \frac{\beta}{2}. \tag{7}$$

3.1.3. Metoda periodogramu

Periodogram $I_N(w)$ jest estymowaną gęstością widmową procesu stochastycznego *X(t)* zdefiniowanego dla przypadków czasu dyskretnego. W celu otrzymania periodogramu badany szereg czasowy jest poddawany operacji dyskretnego przekształcenia Fouriera. Zależność na $I_N(w)$ może być zapisana następująco

$$I_N(w) = \frac{1}{2\pi N}\left|\sum_{k=1}^{N} X_k e^{jkw}\right|^2, \tag{8}$$

gdzie w jest częstotliwością, X_k szeregiem czasowym, a N długością tego szeregu. Własność samopodobieństwa danego procesu wpływa na zachowanie się jego widma $I(w)$ dla $w \to 0$. Zatem dla niskich częstotliwości periodogram może być zapisany w postaci

$$I_N(w) \approx |w|^{-\gamma}, \tag{9}$$

gdzie $0 < \gamma < 1$.

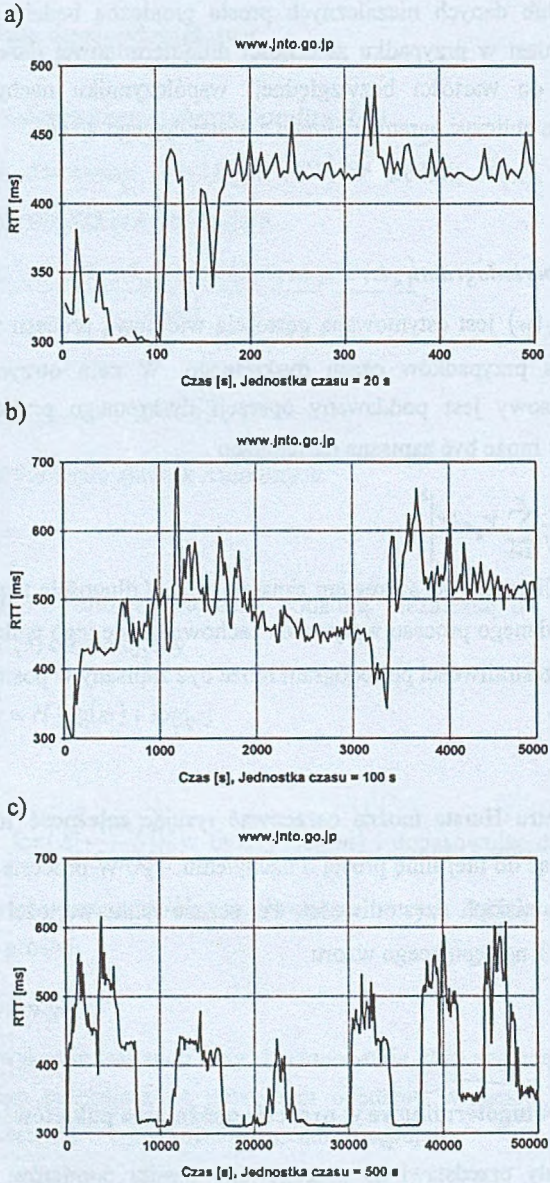
Wartość parametru Hursta można oszacować rysując zależność $\log[I_N(w)]$ w funkcji $\log(w)$ i dopasowując do niej linię prostą o nachyleniu $-\gamma$. W procesie dopasowania należy uwzględniać zakres niskich częstotliwości. Po oszacowaniu wartości γ parametr Hursta może być obliczony z następującego wzoru

$$H = \frac{1+\gamma}{2}. \tag{10}$$

3.2. Zależności długoterminowe w procesie opóźnienia pakietów

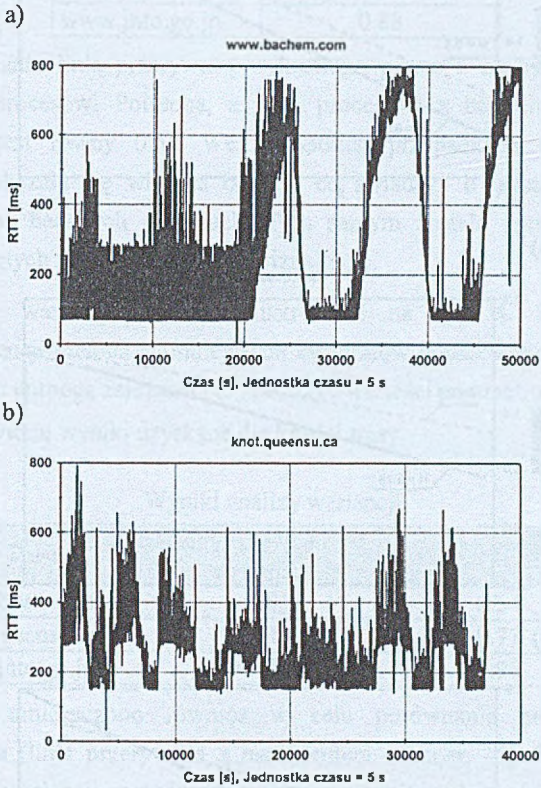
Na rys. 1 zostały przedstawione zagregowane wyniki pomiarów dla pierwszej trasy (*www.jnto.go.jp*) dla trzech różnych rozmiarów okna czasowego (20 s, 100 s, 500 s). Widać, iż zmiany opóźnienia sieci zachowują się podobnie dla małych (minuty, godziny) i dużych (dni) skal czasu. Właściwości statystyczne obserwowanego procesu są zatem niezależne od

skali czasu obserwacji. Takie stwierdzenie uzasadnia postawienie tezy, że obserwowany proces ma charakter samopodobny. Powyższa metoda badania samopodobieństwa jest stosunkowo prosta i opiera się jedynie na intuicyjnej ocenie właściwości szeregu czasowego.



Rys. 1. Round-Trip Time dla trasy www.jnto.go.jp w trzech skalach czasu
Fig. 1. Round-Trip Time for www.jnto.go.jp route over 3 time scales

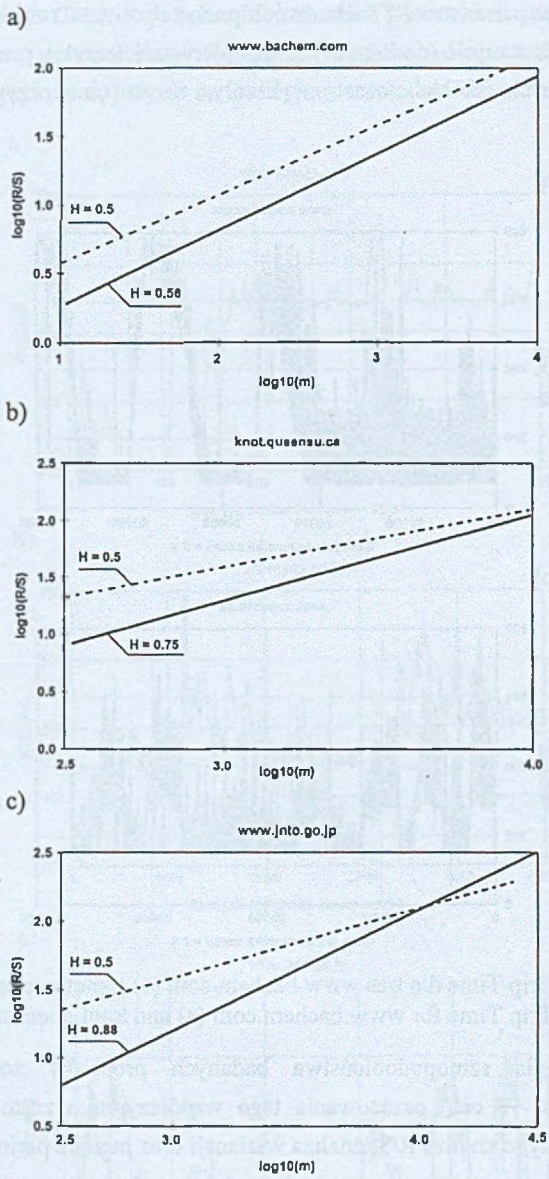
Wyniki pomiarów parametru *RTT* dla dwóch pozostałych tras zostały przedstawione na rys.2. Agregacja tych danych (podobnie jak dla pierwszej trasy) i przeprowadzenie testu wizualnego w różnych skalach czasu wykazały, iż w obu przypadkach występuje samopodobieństwo.



Rys. 2. Round-Trip Time dla tras www.bachem.com (a) i knot.queensu.ca (b)
Fig. 2. Round-Trip Time for www.bachem.com (a) and knot.queensu.ca (b) routes

Do oceny stopnia samopodobieństwa badanych procesów został wykorzystany współczynnik Hursta. W celu oszacowania tego współczynnika zastosowano trzy różne metody graficzne. Były to analiza R/S, analiza wariancji oraz metoda periodogramu.

Na rys. 3 został wykreślony wynik analizy R/S dla trzech tras ruchu. Linią ciągłą zostały narysowane proste dopasowane metodą najmniejszych kwadratów do wykresów funkcji R/S. Nachylenie tych prostych bezpośrednio estymuje wartość parametru Hursta dla danej trasy. Wyniki analizy zamieszczono w tabeli 1.



Rys. 3. Analiza R/S dla tras www.bachem.com (a), knot.queensu.ca (b) i www.jnto.go.jp (c)

Fig. 3. R/S analysis for www.bachem.com (a), knot.queensu.ca (b) and www.jnto.go.jp (c) routes

Tabela 1

Wyniki analizy R/S

Trasa	Parametr Hursta (H)
www.bachem.com	0.56
knot.queensu.ca	0.75
www.jnto.go.jp	0.88

W celu porównania linią przerywaną wykreślono proste z nachyleniem równym 0.5. Odpowiadają one procesowi Poissona, a więc procesowi z brakiem samopodobieństwa (parametr Hursta jest równy 0.5). We wszystkich przypadkach oszacowana wartość parametru Hursta okazała się większa od 0.5, co świadczy o występowaniu zależności długoterminowych w badanych procesach. Tym samym została potwierdzona zasadność wniosków wyciągniętych na podstawie testu wizualnego.

Wyniki analizy wariancji zostały zilustrowane na rys. 4. Do każdej krzywej współczynnika wariancji metodą najmniejszych kwadratów dopasowano proste o nachyleniu $-\beta$. Pozwoliło to za pomocą zależności (7) obliczyć wartości parametru Hursta.
W tabeli 2 przedstawiono wyniki uzyskane dla każdej trasy.

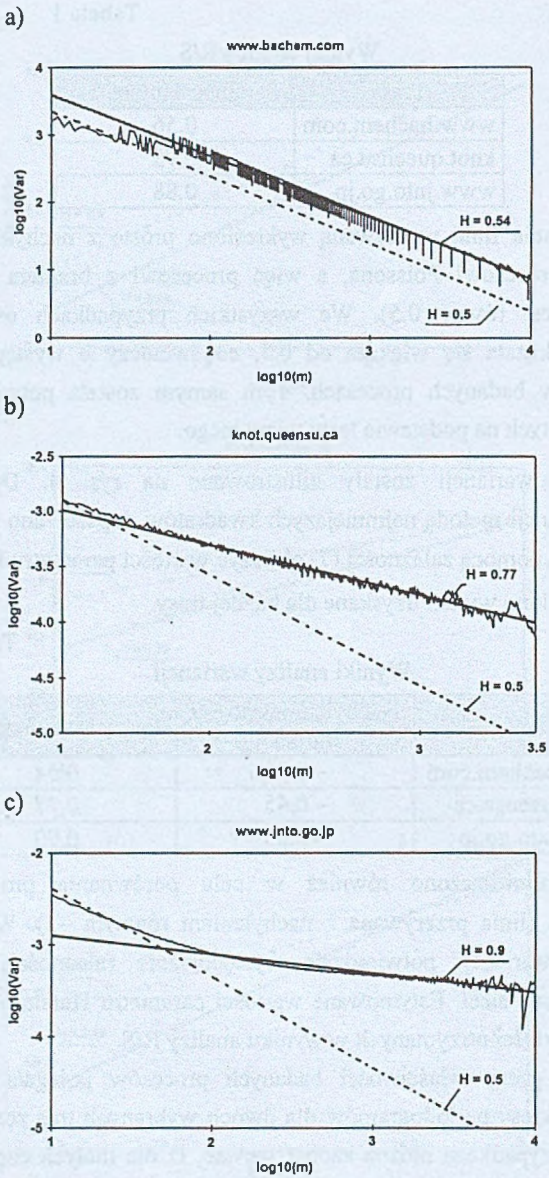
Tabela 2

Wyniki analizy wariancji

Trasa	Nachylenie wykresu wariancji	Parametr Hursta (H)
www.bachem.com	– 0.91	0.54
knot.queensu.ca	– 0.45	0.77
www.jnto.go.jp	– 0.20	0.90

Na wykresach umieszczono również w celu porównania proste odpowiadające rozkładowi Poissona (linia przerywana z nachyleniem równym –1). Wyniki uzyskane na podstawie analizy wariancji potwierdzają występowanie zależności długoterminowych w zmianach opóźnienia sieci. Estymowane wartości parametru Hursta okazały się zbliżone do analogicznych wartości otrzymanych w wyniku analizy R/S.

Trzecia metoda oceny właściwości badanych procesów polegała na wykorzystaniu periodogramów. Wykresy periodogramów dla dwóch wybranych tras zostały przedstawione na rys. 5. W obu przypadkach można zaobserwować, iż dla małych częstotliwości gęstość widmowa jest rozbieżna do nieskończoności. Takie zachowanie gęstości widmowej, charakterystyczne dla szumu 1/f, jest przejawem istnienia zależności długoterminowych. W przypadku procesów z zależnościami krótkoterminowymi wykres gęstości widmowej przebiegałby płasko w całej dziedzinie częstotliwości. Nachylenie wykresów periodogramów $(-\gamma)$ w zakresie niskich częstotliwości zostało oszacowane przez dopasowanie linii prostej metodą najmniejszych kwadratów. Do obliczenia wartości parametru Hursta wykorzystano zależność (10). W tabeli 3 przedstawiono wyniki uzyskane dla każdej trasy.



Rys. 4. Zależność współczynnika wariancji od czasu dla tras www.bachem.com (a), knot.queensu.ca (b) i www.jnto.go.jp (c)

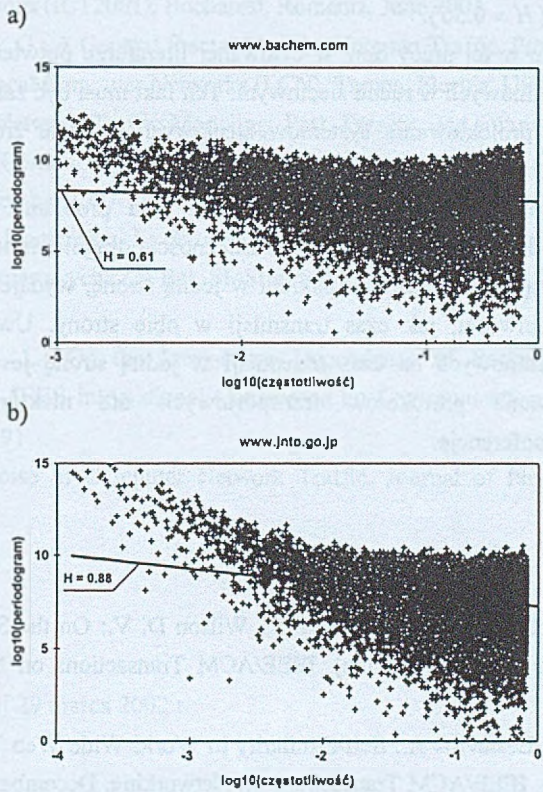
Fig. 4. Variance-time plot for www.bachem.com (a), knot.queensu.ca (b) and www.jnto.go.jp (c) routes

Tabela 3

Wyniki otrzymane metodą periodogramu

Trasa	Nachylenie wykresu periodogramu	Parametr Hursta (H)
www.bachem.com	– 0.22	0.61
www.jnto.go.jp	– 0.77	0.88

Estymowane wartości parametru Hursta otrzymane metodą periodogramu są zgodne z wynikami otrzymanymi innymi metodami i potwierdzają samopodobny charakter zmian opóźnienia badanych sieci.



Rys. 5. Wykresy periodogramów dla tras www.bachem.com (a) i www.jnto.go.jp (b)
Fig. 5. Periodogram plot for www.bachem.com (a) and www.jnto.go.jp (b) routes

4. Wnioski

W pracy dokonano pomiarów szybkości trzech wybranych tras ruchu w sieciach WAN. Analiza statystyczna otrzymanych wyników potwierdziła występowanie zależności długoterminowych w każdym przypadku. Do analizy zostały wykorzystane trzy różne metody estymacji parametru Hursta (analiza R/S, analiza wariancji, metoda periodogramu). Dla każdej trasy wyniki otrzymane za pomocą tych metod były zbliżone. Estymowane wartości parametru Hursta okazały się różne dla różnych tras. Największy stopień samopodobieństwa odpowiadał trasie www.jnto.go.jp ($H \approx 0.88$), zaś najmniejszy trasie www.bachem.com. ($H \approx 0.56$).

Wyniki uzyskane w tej pracy oraz w cytowanej literaturze potwierdzają występowanie zależności długoterminowych w ruchu sieciowym. Ten fakt musi być zatem brany pod uwagę w wielu obszarach projektowania systemów sieciowych (modele źródeł ruchu, kontrola zatłoczeń, wielkości buforów, itp.).

Nasze dalsze badania będą ukierunkowane na dwa problemy. Pierwszy to próba określenia rozkładu długości kolejek w sieci z zależnością długoterminową. Drugi wiąże się z pomiarem i analizą czasu transmisji pakietu w jedną stronę, wydaje się, że powinien on mieć podobne właściwości, jak czas transmisji w obie strony. Uwzględnienie wpływu zależności długoterminowych na czas transmisji w jedną stronę jest szczególnie ważne podczas projektowania protokołów transportowych dla niektórych aplikacji, np. obsługujących telekonferencje.

LITERATURA

1. Willinger W., Leland W. E., Taqqu M. S., Wilson D. V.: On the Self-Similar Nature of Ethernet Traffic (extended version). IEEE/ACM Transactions on Networking, February 1994, s. 1÷15.
2. Crovella M. E., Bestavros A.: Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. IEEE/ACM Transactions on Networking, December 1997, Vol. 5, No. 6, s. 835÷846.
3. Dinh T. D., Molnar S., Vidacs A.: Investigation of Fractal Properties in Data Traffic. Journal on Communications, November-December 1998, Vol. XLIX, s. 12÷18.
4. Czachórski T., Domańska J., Sochan A.: Samopodobny charakter natężenia ruchu w sieciach komputerowych. ZN Pol. Śl. Studia Informatica, Gliwice 2001, Vol. 22, No. 1 (43), s. 93÷108.

5. Park K., Kim G., Crovella M.: On the Effect of Traffic Self-Similarity on Network Performance. Proc. SPIE International Conference on Performance and Control of Network Systems, November 1997.
6. Erramilli A., Narayan O., Willinger W.: Experimental Queueing Analysis with Long-Range Dependent Packet Traffic. IEEE/ACM Transactions on Networking, April 1996, Vol. 4, No. 2, s. 209÷223.
7. Tuan T., Park K.: Congestion Control for Self-Similar Network Traffic. Technical Report CSD-TR-98-014, Purdue University, Dept. of Computer Sciences, May 1998.
8. Popescu A.: Traffic Self-Similarity. Proc. IEEE International Conference on Telecommunications (ICT2001), Bucharest, Romania, June 2001.
9. Molnár S., Terdik G.: A General Fractal Model of Internet Traffic, Proc. 26th Annual IEEE Conference on Local Computer Networks (LCN), Tampa, Florida, USA, November 2001.
10. Ryu B.: Fractal Network Traffic Modeling: Past, Present, and Future. Proc. 35th Allerton Conference on Communication, Control and Computing, Univ. Illinois at Urbaba-Champaign, September 1997.
11. Czachórski T., Domańska J.: Procesy Markowa w modelowaniu natężenia ruchu w sieciach komputerowych. ZN Pol. Śl. Studia Informatica, Gliwice 2001, Vol. 22, No. 1 (43), s. 131÷146.
12. Li Q., Mills D. L.: On the Long-range Dependence of Packet Round-trip Delays in Internet. Proc. IEEE International Conference on Communications, Atlanta GA, June 1998, s. 1185÷1191.
13. Csabai I.: 1/f Noise in Computer Network Traffic. Journal of Physics A, 1994, A27, s. 417÷421.

Recenzent: Dr inż. Mirosław Skrzewski

Wpłynęło do Redakcji 29 marca 2002 r.

Abstract

Many recent studies on network traffic show that the properties of actual traffic are very different from that predicted by traditional teletraffic models. The traffic was described by short-terms dependencies so far. However, it has also long-terms dependencies. It has significant effect on metrics of networking performance, including throughput, packet loss rate, response time and buffer occupancy and it should be take into account by designers of network systems. In this paper selected methods of testing self-similarity processes have been presented. The results of the speed selected routes in WAN have been showed as well. The statistical analysis of measured data showed the existence of long-term dependencies in every case. Three several methods of estimation of Hurst parameter have been used. These methods are R/S statistics analysis (Fig. 3), variance-time analysis (Fig. 4) and periodogram-based analysis (Fig. 5). For given route results obtain with the help these methods were similar. Estimated values of Hurst parameter were different for different routes. The highest level of self-similarity correspond to www.jnto.go.jp route ($H \approx 0.88$) and the lowest correspond to www.bachem.com. route ($H \approx 0.56$). Results obtain in this study as well as in cited papers shows the existence of long-terms dependencies and self-similarity in network traffic. Therefore, this fact must be take into account in many areas of designing of network systems such as traffic models, congestion control, buffer sizing, etc.).